

نشریه دانشکده ادبیات و علوم انسانی

دانشگاه شهید باهنر کرمان

دوره جدید. شماره ۱۹ (پیاپی ۱۶) بهار ۸۵

تاثیر روشهای نوین و چندوجهی ارزیابی بر عملکرد نگارش

دانشجویان در زبان انگلیسی* (علمی - پژوهشی)

اکبر افقري

دانشیار گروه زبان انگلیسی دانشگاه اصفهان

هوشنگ خوش سیمما

استادیار گروه زبان انگلیسی دانشگاه چابهار

چکیده

اخیراً بسیاری از حوزه های تدریس و ارزیابی در رویارویی با فرضیه های جدید، دستخوش دگرگونی شده اند. (بیرنهام، ۱۹۹۶، ص ۲). یکی از این دگرگونی ها، تبدیل روش های سنتی سنجش به روش های نوین ارزیابی است. هدف از این تحقیق قیاس و مقابله روش های سنتی آزمون با روندهای ارزیابی نوین است. در این بررسی، کاستی های روش سنجش سنتی به تصویر کشیده شده است. سپس، به سهمی که روشهای سنجش و ارزیابی در زمینه فراهم آوردن باز خور به دانشجویان داشته اند، اشاره شده است. پیدایش روش ارزیابی چند وجهی و نوین در حوزه مهارت های زبانی و کوشش در زمینه شکل گرفتن تئوری ها و اثر بخشی آنها در زمینه آموزش زبان، مبحث بعدی این مقاله است. در پایان، گزارش نتایج تحقیقی

* تاریخ دریافت مقاله : ۸۴/۵/۲

تاریخ پذیرش نهایی مقاله: ۸۴/۹/۱۴

که توسط نگارندگان انجام شده، آمده است تا نشان دهد که روشهای نوین و چندوجهی ارزیابی بر عملکرد نگارش دانشجویان تاثیر فراوان دارد. واژگان کلیدی: تاثیر گذاری، سنجش، ارزشیابی، ارزیابی نوین و چند وجهی

THE WASHBACK EFFECT OF ALTERNATIVE ASSESSMENT TECHNIQUES ON STUDENTS' WRITING PERFORMANCE

By: Dr.Akbar Afghari, Associate Prof., English Dept. Esfahan
University & Dr. Hooshang Khoshsiman, Assistant Prof. Chabahar
University

Abstract

In recent years, many areas of instruction and evaluation have undergone reforms in response to changing theories of learning, teaching and testing. One such reform is that the era of testing has changed into an era of assessment (Birenbaum, 1996, p.22).

This study aims at comparing and contrasting the traditional methods of testing and the new trends of assessment. In the course of this overview, the traditional testing model and its drawbacks have been drafted. This is followed by dealing with the contributions of testing and assessment to language learning through providing feedback to the students. It continues with the discussion of alternative assessment in the domain of language skills and attempts to form the theoretical basis of the current study on the effectiveness of alternative assessment techniques in language learning. Finally, report of an experiment conducted by the researchers will be given supporting the view that alternative assessment techniques can significantly affect students' writing performance.

**Key words :Wash back effect - measurement – evaluation-
assessment – alternative assessment – formative
assessment- summative assessment- achievement –
ANOVA - t-test - portfolio**

Introduction

The era of testing can be characterized by a separation of instruction and testing activities, by a measurement that was passively undergone by students, and by measuring products solely in the form of a single score. The assessment era promotes integration of assessment and instruction, seeing the student as an active person who shares responsibility, reflects, collaborates and conducts a continuous dialogue with the teacher. Moreover, assessment procedures are seen as valuable for the monitoring of students' progress and directing them, if needed, to remedial learning activities. Hence, the view that the assessment of students' achievement is something that solely happens at the end of a process of learning is no longer tenable (Wolf, 1991, p. 53).

Hancock (1994) claims that assessment is " an ongoing strategy through which student learning is not only monitored but by which students are involved in making decisions about their ability ". He argues that the difference between testing and assessment lies in learners' involvement in the process of making judgments on their own achievement. Additionally, there is a strong support

for representing assessment as a tool for learning. This newly-recognized feature of testing and assessment has persuaded language teachers and testers to employ methods and procedures that enhance students' learning as well as monitoring it.

This trend of shift from testing to assessment and also the move towards employing assessment procedures to enhance learning came into existence following the introduction of learner-centered and communicative teaching methodologies. The reward of such methodologies is a shift from central assessment and central interpretation of assessment results towards the classrooms where assessment occurs for certain specific purposes. *Alternative assessment*, *authentic assessment*, and *classroom assessment* are popular topics which are concerned with employing assessment procedures to raise the standards of assessment and learning. Such assessments introduce several techniques presented in the literature. To name only a few, portfolios, role-plays, oral interviews, self-assessment and conferences are the techniques introduced in such assessment concepts. However, the success of any assessment depends on the effective selection and use of appropriate tools and procedures as well as clear identification of the skill or area to be assessed. Since investigating and implementing all alternative methods were beyond the scope of a single study, certain methods and only one skill have been selected for this study. From among all

the techniques and procedures introduced, only conference and self-assessment were selected as tools and the skill to be assessed was writing.

Significance and Justification of the Study

The field of EFL / ESL has moved from structural teaching approaches to communicative, humanistic and learner-centered approaches. These new approaches in teaching recognize that affective considerations are of vital importance for the acquisition of a foreign/ second language and suggest methods and techniques that create an anxiety-reduced environment for learners (Stevick, 1990, Krashen, 1982, Asher, 1988, cited in Shaaban, 2001, p.16).

Due to the fact that language teaching and testing are closely related, language testing should also enjoy such a shift. The reason is obvious. O'Neil (1992) explains that since new EFL / ESL curricula have moved towards the development of communicative skills, the traditional paper- and- pencil tests are no longer adequate. The traditional summative form of testing which occurs most often at the end of a term of instruction would not be fair to students who are studying on the basis of communicative approach. According to him, there is a need to shift from strictly summative testing tools and procedures to a more humanistic approach using informal assessment techniques that stress formative evaluation which focuses on the processes and products of learning

(p.17).

Such informal assessment techniques also involve students in the process of assessment which consequently improves learning. Brown et al (1997) maintain that the students' involvement in the process of assessment has been proved to be pivotal to effective life-long learning and the development of professional competence (p.16). Since conference and self-assessment are characterized by involving students in the process of assessment, they were used as tools of assessment in this study.

This study intends to find answers to the following question:

What is the washback effect of alternative assessment techniques on EFL students' writing ability?

and it is hypothesized that:

Alternative assessment techniques do not affect students' improvement in writing ability.

Testing vs. Assessment

The importance of testing and assessment in language teaching is well known to all. We often use tests to make decisions about individuals' abilities and our decisions might influence their academic as well as personal lives. Hughes (1989) maintains that information about people's language ability is often very useful and sometimes necessary within teaching systems. He asserts

that as long as it is thought appropriate for individuals to be given a statement of what they have achieved in a second or foreign language, the tests of some kind or other will be needed in order to provide information about the achievement of learners (p.4).

This has made testing and assessment an important component of teaching and instruction. However, care should be taken about using the two terms **testing** and **assessment**. Some applied linguists use the term "**testing**" to apply to the construction and administration of formal or standardized tests such as TOEFL (test of English as a foreign language) and "**assessment**" to refer to more informal methods such as group and peer assessment. For example, Valette (1977) states that "tests" are large-scale proficiency tests and that "assessments" are school-based tests (p.12). This, however, is a rough illustration of the dichotomy between testing and assessment. Bachman (1995) gives a comprehensive definition of testing:

A test is a measurement instrument designed to elicit a specific sample of an individual's performance. As one type of measurement, a test necessarily quantifies characteristics of individuals according to explicit procedures and rules (p.20).

And for Farhady et al. (1994) testing often connotes the presentation of a set of questions to be answered. Assessment, nevertheless, requires a different definition.

According to Shohamy (1992), assessment is a super-ordinate term which includes all forms of assessment. It not only assigns scores to students, but also diagnoses their problems and remedies them through employing specific methods and techniques (p.54). Gipps (1994) also defines assessment as "a wide range of methods for evaluating pupils' performance and attainment" including formal testing and examinations, practical and oral assessment and classroom-based assessment carried out by teachers (p.10).

Regarding the importance of assessment in contrast to testing, Inger (1993) argues that testing is designed to be administered during a normal school period and it presents a series of discrete tasks that force students to move repeatedly from one unconnected item into the next. Inger concludes that this shortcoming of language testing can be overcome by assessment techniques and procedures (P. 7).

Traditional Tests

Traditional tests are based on psychometrics principles which developed from work on intelligence and intelligence testing. Cunningham (1998) calls traditional paper-and-pencil techniques as conventional tests and divides them into two main categories:

1. Objective or selected response items that require students to choose the answer from among several

choices like true-false, multiple-choice and matching exercises; and

2. constructed response items that require students to create their own response. These include short answer, completion, fill-in-the-blank and essay-type items (p.123).

Cunningham further states that conventional assessment techniques are based on psychometric assumptions. These assumptions include:

1. The responses of students on the test are the same as physical objects which can be counted, turned into numbers and interpreted as statistics;
2. Testing requires the use of a neutral, value-free scientific language to describe student performance; and
3. Testing is objective and takes place independently of its context (p.124).

Gipps (1994) also asserts that in the traditional testing model one can specify and measure all important learning objectives, and furthermore, mastery on the test items implies mastery of the intended skills and concepts (p.9).

However, with the advent of communicative teaching methodology and learner-centered approaches, traditional tests have been brought into question regarding their

validity in assessing real-life tasks. The discrete-point testing seems no longer adequate. From the majority of scholars' point of view, traditional and objective tests suffer from several disadvantages and drawbacks. Carrol (1961), for example, was amongst the first people who criticized the testing of language bit by bit and suggested using tests which encompasses all the components of language simultaneously. The two main categories of traditional tests are briefly discussed here with regard to their disadvantages:

Selected-response Tests

Selected-response tests include true-false, matching and multiple-choice items. Brown and Hudson (1999) consider two disadvantages for the selected-response tests as follows:

- a. They are relatively difficult for the test writer to construct, and
- b. They do not require students to use any productive language (p.685).

Multiple-choice tests, as one major and quite commonly-used selected-response testing technique, have been criticized by Hughes (1989):

- a. The technique [multiple-choice test] tests only recognition knowledge;
- b. Guessing may have a considerable but unknowable

- effect on test scores;
- c. The technique severely restricts what can be tested;
 - d. It is very difficult to write successful items;
 - e. Washback effect may be harmful; and
 - f. Cheating may be facilitated (pp. 60-62).

Along the same line, Heaton (1988) states that "the chief criticism of the multiple-choice testing is that frequently it does not lend itself to the testing of language as communication" (p.27).

The use of true-false items, also, is not recommended. According to Farhady et al. (1994):

True-false items, though frequently used in language tests, are not highly recommended because of two reasons: first, they very much depend on chance, namely, the examinee has a fifty percent chance of getting a correct response without having any knowledge of the points being tested. Second, they are limited to measuring simple learning activities in language. Complex tasks cannot be measured validly through true-false items. These two shortcomings decrease the reliability, validity and application of true-false items (pp 89-90).

Nevertheless, as Heaton (1988) states, selected-response items and specifically multiple-choice items have some advantages too: they can offer a useful introduction to the construction of objective tests... . They are advantageous in measuring students' ability to recognize

correct grammatical forms, etc, and to make important discrimination in the target language. On the whole, multiple-choice items can help both students and teacher to identify areas of difficulty (p. 27).

Constructed-response tests

These items include fill-in-the-gap and short-answer forms. It is stated by Brown and Hudson (1999) that a fill-in-the-gap test normally focuses on testing a single word or short phrase at most. Another problem is that a blank to fill in may have a number of possible answers. Then, Brown and Hudson argue that short-answer tests focus on testing a word or a phrase. A second disadvantage is that multiple answers are possible, which means that if the problems are not carefully stated, each student may produce a completely different answer (p. 661).

However, constructed-response items have also some advantages. In general, they have no guessing factor and they measure productive language use as well as the interaction of receptive and productive skills (p.662).

New Trends

within the last two decades a new generation of testing came into existence which had a profound effect on the principles and procedures of language testing. This new generation is the ***personal-response assessments*** and includes portfolio assessment, self-assessment and so on.

Brown and Hudson (1999), considering the two aforementioned categories of first generation, comment on the second generation of tests as:

In general, personal-response assessments are beneficial in that they provide personal or individualized assessment, can be directly related to and integrated into the curriculum, and can assess learning processes in an ongoing manner throughout the term of instruction (p.663).

However, in the relevant literature of recent decades, the personal-response assessments are referred to as alternative assessments. This reform in assessment is assumed to enhance the effect of testing techniques on teaching and learning. In fact, the move to reform assessment is based upon the premise that assessment should primarily support learning and this is achieved through providing students with the positive feedback. Due to the significant role of assessment and feedback in enhancing learning, the relationship between assessment, learning and feedback is discussed in the next section.

Assessment, Learning and Feedback

It is widely accepted that testing has an effect on teaching. What is not so often acknowledged is the relationship between assessment and what and how pupils learn (Gipps, 1994, p.18).

Glaser (1994) points out that " as assessment and

instruction are more closely linked, achievement measurement will be integral to learning rather than imposed by some external examination on students' fates " (p.26). Then he emphasizes the significant role of assessment in learning and states:

“The nature of assessment will necessitate analysis of the cognitive aspects of a task and the performance that it entails. The closer ties between assessment and instruction imply that those performances will become more apparent to students and teachers... the performance criteria by which students are judged will be evident so the criteria can motivate and direct the process of learning “ (p.27).

He further claims:

“As performance criteria become more openly available, students will become better able to judge their own performance without necessary reference to the judgment of others. Instructional and assessment situations will provide coaching and practice in ways that help students reflect on their performance. Occasions for self-assessment will enable students to set incremental standards by which they can judge their own achievement and develop self-direction for attaining higher performance levels...” (pp. 27-28).

In addition, regarding the role of assessment in learning and instruction, Glaser maintain that if teachers use new

forms of assessment to improve their teaching and if they, together with educational policy makers, devise systematic approaches to integrate assessment into learning and instruction, perhaps the time for change in assessment will be at last upon us (p. 28).

Assessment can also be employed to influence students' performance. As Stephan(1994) claims, if assessment procedures are chosen correctly, assessment can develop and facilitate students' performance (p. 4). However, in order for the assessment to facilitate students' performance and consequently to improve their learning, it should benefit from positive **feedback**: Richards at al (1992) define feedback as " comments or information learners receive on the success of a learning task, either from the teacher or from other learners " (p.137). Black and William (1998) also emphasize the role of feedback in learning and elaborate on the ways feedback can be made effective for students' learning:

Research studies have shown that if pupils are given only marks or grades, they do not benefit from the feedback. The worst scenario is one in which some pupils who get low marks this time and come to expect to get low marks next time... Feedback has shown to improve learning when it gives each pupil specific guidance on strengths and weaknesses, preferably without any overall marks...Pupils must be given the means and opportunities to work with evidence of their difficulties. For formative

purposes, a test at the end of a unit or teaching module is pointless; it is too late to work with the results. We conclude that the feedback on tests, seatwork, and homework should give each pupil guidance on how to improve, and each pupil must be given help and an opportunity to work on the improvement.

Chastain (1988) also comments on utilizing feedback in the process of assessment and maintains that students should know which goals they have failed to achieve and which weaknesses in their preparation they should try to remedy. He believes that if no feedback is given to the students and if the results of their efforts are not appraised, many of them will not be able to monitor their progress adequately. He continues that this process of feedback involves both the teacher and students in the trend of assessment rather than overestimating one's role to the exclusion of the other's (p. 394).

The feedback provided in the process of assessment has some uses for both teachers and students. Sadler (1989) considers feedback as essential component of every teaching – learning process and states that teachers use feedback to make programmatic decisions with respect to readiness, diagnosis and remediation. Students use it to monitor the strengths and weaknesses of their performances, so that aspects associated with success or high quality can be recognized and reinforced, and unsatisfactory aspects modified or improved.

Sadler believes that feedback from the teacher needs to be of the kind that helps the students in comparing the actual performance with the desired performance and tells them what to do to improve. As he believes, the use of grades or "good, 7/10" marking cannot do this. Information fed back to the students is only feedback when it can be used to close the gap between the actual performance and the desired performance (cited in Gipps, 1994, p. 125).

The importance of utilizing feedback in teaching / learning process has some grounding. As Gipps (1994) believes, feedback, in the process of teaching, is considered to be important for two reasons: it contributes directly to progress in learning through the process of formative assessment, and indirectly through its effect on pupil's academic self-esteem. Thus, as this brief review reveals, feedback has long been recognized as an important feature of the teaching – learning process. The model by Bennett (1982), for example, includes teacher feedback which is regarded as crucial for both pupil involvement in learning and hence achievement. Bennett considers feedback to be one of the structuring conditions for learning (cited in Gipps, 1994, p.130).

The literature on the impact of feedback is rich with studies which prove that positive feedback can enhance learning. In his review of research, Crooks (1988) reports on the impact of classroom evaluation on

students' performance and concludes that feedback assists learning. He suggests that teachers need to make more use of learning-related feedback and less use of feedback for evaluation or grading purposes (p.17). Also, a meta-analysis by Kulik, Kulik and Bangert-Drowns (1991) supported the effect of feedback and remediation on improving learning at all levels of schooling.

Regarding the important role of feedback in teaching and learning over the past several years, a contrary view has emerged and gained acceptance which has recognized that testing in its traditional form fails to provide positive feedback for the instruction and teaching and therefore does not enhance learning. During the 1990s, new forms of educational assessment rose to prominence and it became necessary to distinguish these techniques from those that had dominated testing since its inception. As mentioned before, these new forms are called **alternative assessments**.

Alternative Assessments

Various descriptions of alternative assessment exist in the literature. Frechtling (1991), among others, sees alternative assessment as a means of escaping the problems of multiple-choice testing. He believes that the new methods go beyond simplistic multiple-choice questions and require students to perform in situations

that are both more life-like and more complex (cited in Cunningham, 1998, pp.120-124).

Alternative assessment techniques have received much attention in the last decade and several forms of assessment have been introduced recently. Huerta-Macias (1995) gives a list of alternative methods including checklists, journals, videotapes and audiotapes, and teacher observation (p.9). Some other types are portfolios, presentations, written narratives, oral interviews, role-plays, student-teacher conferences, and self- and peer- assessments (Shaaban, 2001, pp.19-21).

Among the various types of alternative assessment, self assessment and portfolio assessment which have been utilized in this study are explained below.

Self-Assessment

The increased interest in involving the learner in all phases of the learning process and in encouraging learner autonomy and decision making has led to the interest in self-assessment (Alderson, et al, 1993, p.227). According to Brown and Hudson (1999), self-assessment requires students to rate their own language. They count a number of advantages for self-assessment: first, self-assessment can be designed to be administered relatively quickly. Second, they inevitably involve students directly in the assessment process. Third, such involvement may help students understand what it means to learn a

language autonomously. Finally, both the students' involvement and their greater autonomy can substantially increase their motivation to learn the language in question (p. 666).

In fact, self-assessment is used to relegate more responsibility to students to identify where they believe they have been successful and where they believe they require assistance. It refers to the involvement of learners in making judgments about their own learning, particularly about their achievements and the outcomes of their learning. It is a way of increasing the role of students as active participants in their own learning and is mostly used for formative assessment in order to foster reflection on one's own learning process and results (Boud and Brew, 1995, p.9). Klenowsky (1998) considers self-assessment as a cognitive strategy which is the main reason for the paradigmatic shift in assessment where instead of simply measuring the learning, the learning itself is focused on (p.21).

Portfolio Assessment

According to Brow (1998), the concept of portfolio was borrowed from the field of fine arts where portfolios are used to display the best samples of an artist's work. Portfolio is a systematic collection of a variety of teacher observations and student products, collected over time, that reflect a student's developmental status and progress. Portfolio is not a random collection of

observations or student products; it is systematic in that the observations that are noted and the student products that are included relate to major instructional goals (cited in Shabban, 2001, p.30).

Genese and Upshur (1996) maintain that the primary value of portfolios is in the assessment of student achievement, because they provide a continuous record of students' language development that can be shared with others. They further state that portfolios can increase the students' involvement in and ownership of their own learning (p. 99). The positive effects of portfolios on student learning is why their use is highly encouraged in the literature on alternative assessment.

The Rationale for Alternative Assessment

Over the past several years, there has been a great interest in using alternative assessment techniques. The interest and support has been justified in the literature by different scholars. Cunningham (1998) for example, considers some reasons for the remarkable support for alternative assessment:

1. concern about the negative impact of the use of standardized tests in minimum competency testing;
2. dissatisfaction with existing psychometric models;
3. the belief that the primary purpose of public schools

is the promotion of social justice (p.124).

Cunningham, then, asserts that the use of alternative assessment is based on assumptions about how students learn; how best to teach them; and the role of assessment, that are quite different from conventional assumptions in these areas. He continues that alternative assessment can be viewed as a rejection of conventional principles of measurement, educational testing and instruction. Conventional methods are indirect and artificial and teachers face many difficulties trying to prepare students for such tests (p. 128).

Moreover, it should be mentioned that most alternative techniques emphasize formative assessment and they can help decrease the level of anxiety caused by concentration on linguistic accuracy and since they stress communicative fluency, they can increase students' comfort and feeling of success (Shaaban, 2001, p.18).

Hancock (1994) believes that in the real world most of us have more than one opportunity to demonstrate that we can complete tasks successfully. So it makes sense to provide similar opportunities for students in instruction and assessment. It means that meaningful authentic assessment should be used that involves language learners in the process of assessment and offers teachers a wide range of evidence on which to judge whether students are becoming competent and purposeful language learners (p. 11).

Finally, Heurta-Macias (1995) offers several characteristics of alternative assessment. According to him, they are:

1. non-intrusive in that they extend the day-to-day classroom activities already in a curriculum;
2. allow students to be assessed on what they normally do in class;
3. provide information about both the strengths and weaknesses of students; and
4. are multi-culturally sensitive when properly administered.

Aschbacher (1991) lists other characteristics of alternative assessments as follows:

1. They require problem-solving and higher level thinking.
2. They involve tasks that are worthwhile as instructional activities.
3. They use real-world contexts or simulations.
4. They focus on processes as well as products; and
5. they encourage public disclosure of standards and criteria (cited in Brown and Hudson, 1999, p. 653).

Advantages of alternative assessment techniques over the

traditional summative models, require that they should be integrated into the present day practices of Teaching English as a Foreign Language (TEFL).

Alternative assessment includes a variety of measures that are suited for assessing different language skills. However, no single assessment model is suited for every purpose. The issue is not whether one form of assessment is intrinsically better than another, rather, the nature of some techniques are such that they can lend themselves to some skills more than the others. For example, as the literature on the alternative assessment techniques shows, portfolio assessment is used to assess students' reading, writing, and listening skills; journals to assess writing skill; interviews and role-plays to assess speaking skill and written narratives to assess writing skill only.

However, there are some techniques which lend themselves to all language skills. For the purpose of this study, *the two techniques of **portfolio** and **self-assessment*** were selected to see if they have any effect on students' writing skill ability.

Method

Participants

A group of 60 students majoring in English from the Maritime university of Chabahar participated in this

study. 30 students served as experimental group and 30 students as control group. They attended four-hour-a-week writing classes.

Instruments

1. A general proficiency test was validated and administered to both groups as the pretest to check the homogeneity of the two groups. The reliability coefficient of pretest was .82.

2. A parallel general proficiency test was validated and administered to both groups as posttest 2 to see the potential differences between the performance of the two groups. The reliability coefficient of posttest 2 was .79.

3. An IELTS Academic Writing Test was also administered to both groups as posttest to check the potential differences in writing performance of the two groups. This test which was an integrative reading and writing test included 5 items to recognize irrelevant sentences, 5 items to put sentences in order, 5 items to use cohesive devices, 5 items to determine the function of sentences and a writing task using a graph.

Procedures

At the outset of the semester, the pretest was administered to all participants. The purpose was to ensure the homogeneity of the two groups with respect

to their English language proficiency.

Throughout the three-month semester, the two alternative assessment techniques were utilized for the experimental group, based on the activities of the textbook. The procedure for implementing these two techniques was as follows:

In each session, students were asked to write a written narrative which is one of the assessment techniques. Then the teacher observed their assignments and gave them some guidelines to review their narrations. Certain criteria were given to students to help them judge their own tasks. Using students own narrations as samples, relevance, appropriate organization, sentence variety, and cohesion were introduced to them and they were asked to assess their task based on these criteria.

Based on the information provided by the participants on their own strengths and weaknesses, the instructor grouped the participants according to the problems they had in writing: those who had problems with coherence ; those who had problems with organization; those who had difficulty with structures; those who had difficulty with cohesion; and finally, those who had problems with sentence variety. Accordingly, in the second session, the instructor provided feedback to each group to help them overcome their weaknesses.

The participants in the experimental group had 10 tasks to do

during the semester. Each task was dealt in two sessions simultaneously through observations and records made by the instructor relevant feedback was provided for ratings.

In the comparison group, the routine syllabus was followed without any resort to alternative techniques, i.e. there was no self assessments and portfolios. They wrote essays in each session and these essays were scored by the instructor without involving students in the process of assessment.

At the end of the course, all participants in the two groups took a traditional standard proficiency test as well as an IELTS writing qualification test which was rated based on the principles of writing paragraphs/essays and the predetermined criteria. These tests served as the posttests to investigate the effect of the treatments on the experimental group's achievements and learning during the period of instruction.

Finally, the experimental and the control group were compared on the basis of their mean scores for research interpretation.

Design of the Study

This study involved one independent variable and one dependent variable. The independent variable was the

alternative assessment techniques, including self-assessment and portfolio. The dependent variable was the scores of the participants on portfolio and self-assessment, general proficiency and writing performance posttests.

Regarding the nature of the research question and hypothesis, the most appropriate design was intact group design which is a sub-category of pre-experimental method of research.

Data Analysis Techniques

In order to test the Research Hypothesis, the following statistical techniques were utilized:

1. To check the homogeneity of the two groups at the outset of the experiment, an Independent t-test was performed.
2. To check the degree of relationship between all the variables, a Pearson correlation was run.
3. To see the difference in the performance of the participants within each group on two posttests, two matched t-tests were conducted.
4. To check if students in the two groups made any improvement from pretest to posttests, the gain scores were calculated for both groups:

Then, to see the difference between the two groups in terms of their gain scores, two Independent t-tests were performed.

5. To understand the difference in the performance of the participants within the experimental group on 10 self-assessments, a repeated-measure ANOVA was run.

Findings

To check the homogeneity of the two groups, they were pretested through a general proficiency test. The descriptive statistics of the pretest is reported in table 1.

Table1 Descriptive Statistics

Groups	N	Mean	Std. Deviation
Experimenta	30	12.60	3.14
Comparison	30	10.06	4.45

As the table shows, the mean of the two groups is almost different. This is true concerning the standard deviation of the two groups. However, an Independent t-test was run to see if the observed difference is statistically significant. Table 2 presents the results.

Table 2. T- test Analysis of Pretest Scores

	Df	Mean Difference	T	Sig
Groups	58	2.53	2.45	.024

significant difference between the two groups in terms of their performance on the pretest at the outset of the study is revealed by the t-ratio(sig= .01).

The second set of analyses included the degree of go-togetherness of the scores. The correlation coefficients (sig = .05) are presented in table 3.

Table 3. Correlation Matrix

	Pre	Portfo 1	portfo 2	portfo 3	Portfo 4	portfo 5	portfo 6	portfo 7	portfo 8	portfo 9	portfo 10	Post 1	Post 2
Pre	1.00												
Portfo. 1	.15	1.00											
Portfo. 2	.27	.59	1.00										
Portfo. 3	.27	.56	.72	1.00									
Portfo. 4	.27	.66	.90	.86	1.00								
Portfo. 5	.28	.40	.73	.54	.73	1.00							
Portfo. 6	.17	.48	.74	.73	.78	.56	1.00						
Portfo. 7	.26	.51	.76	.56	.72	.75	.60	1.00					
Portfo. 8	.22	.56	.74	.68	.82	.63	.63	.69	1.00				
Portfo. 9	.21	.37	.65	.68	.73	.51	.56	.61	.71	1.00			
Portfo.10	.23	.62	.65	.69	.78	.65	.69	.58	.78	.57	1.00		
Post 1	.45	.41	.53	.56	.61	.52	.52	.51	.50	.49	.50	1.00	
Post 2	.54	.36	.43	.47	.49	.36	.52	.33	.39	.35	.44	.50	1.00

Based on the data reported in Table 3, it can be assumed that there is a high correlation between all portfolio assessment scores whereas there is a low correlation between pretest, posttest 1 and posttest 2. The high correlation between portfolios can be attributed to the similar nature of these assessments and the low correlation between pre and posttests, especially between pre and posttest 1 and between posttest 1 and posttest 2, can be attributed to the different nature of these tests.

The third set included two matched t-tests to see if there is a difference between the performance of the participants in each group on posttests. The reason for conducting these two analyses was the different nature of posttests; posttest 1 being a qualification test and posttest 2 a discrete-point traditional test. Consequently, the first t-test was run on the performance of the experimental group on the two posttests. The results are reported in table 4.

Table 4 Matched t-test for Experimental Group

Group	df	Mean	Std. Deviation	Mean Difference	t	Sig
Experimental	29	16.90	2.72	3.50	4.48	.000
		13.40	2.98			

As the t-ratio (sig = .000) reveals, there is a significant difference in the performance of the experimental group

on posttest 1 and posttest 2.

The second set of matched t-test was run on the performance of the control group on the two posttests. Table 5 shows the result.

Table 5 Matched t-test for Comparison Group

Group	df	Mean	Std. Deviation	Mean Difference	t	Sig
Comparison	29	9.23	4.88	1.20	1.06	.29
		8.03	5.73			

As the t-ratio (sig = .29) shows, there is no significant difference in the performance of the comparison group on the two posttests. Thus, it can be safely concluded that due to the treatment effect, the experimental group performed better on the communicative test while the comparison group made no difference in their performance on the communicative vs. traditional test.

To check whether students in the two groups had any improvements from pretest to posttests, their gain scores were calculated. To see the difference between the gains of the two groups, the fourth set of analyses was performed which included two independent t-tests on two sets of gain scores. Table 6 shows the result.

Table 6. T-test Analysis of (pretest – posttest 1)

Groups	Mean	SD	df	Mean difference	T	sig
Experimental	4.30	4.66	58	5.13	3.93	.000
Comparison	-.83	5.41	58			

As the t-ratio (sig = .000) shows, there is a significant difference between the two groups in terms of their gain scores from pretest to posttest 1 which was communicative in nature. Table 7 reports the result of the second t-test.

Table 7. T-test Analysis of (pretest – posttest2)

Groups	Mean	SD	df	Mean difference	T	Sig
Experimental	.76	4.40	58	1.80	1.23	.22
Comparison	-1.03	6.69	58			

The t-ratio (sig = .22) reveals that there is virtually no significant difference between the gains experimental and comparison groups made from pretest to posttest 2 which was traditional in nature.

The results of the two t-tests, conducted on the gain scores, indicates that employing alternative assessment

techniques in instruction can lead to a better performance on communicative rather than traditional tests.

The fifth set of analyses included a repeated-measure ANOVA on all portfolio/self-assessment scores of the experimental group. The basic concern was to see if the students' performance differed on 10 portfolio/ self-assessments. The result are reported in table 8.

Table 8- Repeated-measure ANOVA on portfolio/ self-assessments

Source of Variance	Type III Sum of Squares	df	Mean Square	F	Sig.
SELF (within group)	33.204	9	3.689	3.808	.000
Error	252.895	261	969		

As the F-ratio reveals, there is a significant difference in the performance of the experimental group on 10 portfolio/self-assessments.

Conclusion

Based on the findings of the research, the effectiveness of the treatment was confirmed. It can be concluded that employing alternative assessment techniques in the

assessment of writing skill lead to a significant difference between the performance of the two groups. The findings of this study are in agreement with the existing studies in the literature which reveal that alternative assessment methods and procedures can enhance students' achievements. The findings are also in line with the findings of all the studies showing the usefulness of involving students in the process of their own assessment as well as providing feedback in instruction.

Based on the results, it can be concluded that the experimental group performed better due to the alternative assessments utilized for them as the treatment. As the experimental group had a better performance on the qualification rather than traditional posttest, it can be justified that alternative assessment techniques are in line with the reform and the shift from testing to assessment in language education.

References

- Alderson, J. C. & Wall, D. (1993). 'Examining Washback: the Srilankan Impact study'. *Language Testing*, 10, 1, 120.
- Birenbaum, M. (1996). 'Assessment 2000: towards a pluralistic Approach to assessment', in: M. Birenbaum & F. Dochy(EDS). *Alternatives in Assessment of Achievement, Learning, Processes and Prior Knowledge*, pp. 3-31 Boston, Kluwer Academic.
- Black, P. & William, D. (1998). 'Assessment and Classroom Learning', *Assessment in Education*, 5, 1: 7-74.
- Boud, D. & Brew. A. (1995). *Enhancing Learning through Self-Assessment*. London and Philadelphia, Kogan Page.
- Brown J. D & Brown, S. (1997). *Assessing Student Learning in Higher Education*. London and Newyork: Routledge.
- Brown S. & Hudson, T. (1999). *The Alternatives in Language Assessment*. *TESOL Quarterly*, 32, 4, 653-673.
- Carrol, J. B. (1961). *Fundamental Considerations in Testing for English Language Proficiency of Foreign Students*. Washington, D. C: Routledge.
- Chastain, K. (1988). *Developing Second-Language Skills: Theory and Practice*. Florida: Harcourt Brace Jovanovich, Inc.

Crooks, T. J. (1988). ' *The Impact of Classroom Evaluation Practices on Students* ' Review of Educational Research, 58, 4:38,39.

Cunningham, G. K. (1998). *Assessment in the Classroom*. London: Falmer Press.

Farhady, Jafarpu, and Birjandy (1994). Testing Language Skills from Theory to Practice. Tehran: SAMT.

Genesee, F. & Upshur, J. (1996). *Classroom-based Evaluation in Second Language Education*. Cambridge: CUP.

Gipps, C. V. (1994). *Beyond Testing: Towards a Theory of Educational Assessment*. London: Falmer Press.

Glaser, R. (1990). ' *Towards New Models for Assessment* ', International Journal of Educational Research, 14, 5, 475-83.

Hancock, C. R. (1994). *Alternative Assessment and Second Language Study: What and Why?*, The Ohio State University, [on-line]. Retrieved February, 2003. Available: www.ed.gov.databases/ERIC_Digests/ed376695.html- 22k.

Heaton, J. B. (1988). *Writing English Language Tests*

(New Ed). London: Longman.

Hughes, A. (1989). *Testing for Language Teachers*. Cambridge: CUP.

Huerta-Macias, A. (1995). *Alternative Assessment: Responses to Commonly Asked Questions*. TESOL Journal, 4, 4, 8-10.

Inger, M. (1993). *Authentic Assessment in Secondary Education*, IEE BRIEF [on-line]. Retrieved June, 2004. Available: www.ERIC_Digests.com.

Klenowsky, V. (1998). '*Connecting Assessment and Learning: Portfolio Assessment*', A Paper Presented in a Conference held in China.

Kulik, C. C. and Halley, F. S. (1991). '*Effects of Frequent Classroom Testing*', Journal of Educational Research, 85, 1: 81-99.

Richards et al. (1992). Longman Dictionary of Language Teaching & Applied Linguistics (2nd Ed). Great Britain, London.

Shaaban, K. (2001). *Assessment of Young Learners*. English Teaching Forum, 39, 4, 16-22.

Shohamy, E. (1992). '*New Modes of Assessment: The Connection Between Testing and Learning*', in: E. Shohamy & R. Walton (Eds). *Language Assessment for Feedback: Testing and Other Strategies*. Dubuque, Iowa: Kendall Hunt Publishing Company.

Stephan, N. E. (1994). *Creating Meaningful Performance Assessment*. London: OUP.

Tannenbaum, J. E. (1996). *Practical Ideas on Alternative Assessment for ESL Students*, Montgomery County Public School [on-line]. Retrieved February, 2003.

Valette, R. M. (1977). *Modern language Testing*. (2nd Ed) . New York: Harcourt Brace Jovanovich.

Wolf, G. (1991). 'To Use Their Minds Well: *Investigating New Forms of Student Assessment*', Review of Research in Education, 17, pp. 31-74.